

Bounding The Rate of Adaptation In A Large Asexually Reproducing Population With Fast Mutation Rates

by Michael Kelly
 University of California, San Diego

August 3, 2011

Abstract

We consider a model of asexually reproducing individuals. The birth and death rates of the individuals are affected by a fitness parameter. The rate of mutations that cause the fitnesses to change is proportional to the population size, N . The mutations may be either beneficial or deleterious. In a paper by Yu, Etheridge and Cuthbertson (2009) it was shown that the average rate at which the mean fitness increases in this model is bounded below by $\log^{1-\delta} N$ for any $\delta > 0$. We achieve an upper bound on the average rate at which the mean fitness increases of $O(\log N / \log \log N)$.

Keywords: Evolutionary process, Moran Model, Selection, Adaptation rate

1 Introduction

In a finite, asexually reproducing population with mutations it is well-known that competition among multiple individuals that get beneficial mutations can slow the rate of adaptation. This phenomenon is known as the Hill-Robertson effect, named for the authors of [7]. One may wish to consider the effect on the rate of adaptation of a population when there are many beneficial mutations present simultaneously. It is easily observed that when such a population is finite and all mutations are either neutral or deleterious then the fitness of the population will decrease over time. This scenario is known as Muller's ratchet. The first rigorous results regarding Muller's ratchet were due to Haigh [6]. In an asexually reproducing population beneficial mutations are necessary to overcome Muller's ratchet. Yu, Etheridge and Cuthbertson [11] proposed a model that gives insight into both the Hill-Robertson effect and Muller's ratchet in large populations with fast mutation rates.

The model introduced in [11] is a Moran model with mutations and selection. In our model there is a population of N cells and positive constants μ , q and γ which do not depend on N and determine the dynamics of the system. The parameters must satisfy $\mu > 0$, $0 < q \leq 1$ and $\gamma > 0$.

Contact: mbkelly@math.ucsd.edu, Mathematics Department, UC San Diego, 9500 Gilman Dr. #0112, La Jolla, CA 92093, U.S.A.

AMS 2010 subject classifications. Primary 92D15; Secondary 60J27, 82C22, 92D10

Let X_t^i be the fitness of cell i at time t . Then $X = (X^1, X^2, \dots, X^N)$ is a stochastic process with state space \mathbb{Z}^N . The system has the following dynamics:

1. Mutation: Each individual acquires mutations at rate μ . When cell i gets a mutation it is beneficial with probability q and X^i increases by 1. With probability $1 - q$ the mutation is deleterious and X^i decreases by 1.
2. Selection: For each pair of individuals (i, j) , at rate $\frac{\gamma}{N}(X^i - X^j)^+$ we set X^j equal to X^i .
3. Resampling: For each pair of individuals (i, j) , at rate $1/N$ we set X^j equal to X^i .

Our model only differs from the one in [11] by allowing $q = 1$. The upper bound we establish for the rate of adaptation still holds in the absence of deleterious mutations.

Under the selection mechanism the event that X^j is set to equal X^i represents the more fit cell i giving birth and the less fit cell j dying. Likewise, the resampling event that causes X^j to equal X^i represents cell i giving birth and cell j dying.

A heuristic argument in [11] shows that as N tends to infinity the mean rate of increase of the average fitness of the cells in X is $O(\bar{k})$ where \bar{k} is the number satisfying

$$\bar{k}^{\bar{k}} = N.$$

We note that this is equivalent to being $O(\log N / \log \log N)$. The heuristic argument is difficult to extend to a rigorous argument. Let

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X^i$$

be the continuous-time process which represents the average fitness of the cells in X and start X as centered about 0 so that $\bar{X}_0 = 0$. The rigorous results established in [11] are as follows:

- The centered process X^C , in which cell i has fitness $X^{C,i} = X^i - \bar{X}$, is ergodic and has a stationary distribution π .
- If

$$c_2 = E^\pi \left[\frac{1}{N} \sum_{i=1}^N (X^{C,i})^2 \right]$$

is the variance of the centered process under the stationary distribution, then

$$E^\pi[X_t] = (\mu(2q - 1) + \gamma E^\pi[c_2])t,$$

where E^π means that the initial configuration of X is chosen according to the stationary distribution π .

- For any $\delta > 0$ there exists N_0 large enough so that if $N \geq N_0$ then $E^\pi[\bar{X}_1] \geq \log^{1-\delta} N$.

Rigorous estimates for $E^\pi[c_2]$ are difficult to find so other methods are needed to compute $E[\bar{X}_t]$. The third result of [11] shows that if there is a positive ratio of beneficial mutations then a large enough population will increase in fitness over time. A paper by Etheridge and Yu [5] provides further results pertaining to this model.

Other similar models can be found in the biological literature. In these models the density of the particles is assumed to act as a traveling wave in time. The bulk of the wave behaves approximately deterministically and the random noise comes from the most fit classes of cells. One tries to determine how quickly the fittest classes advance and pull the wave forward. This traveling wave approach is used in [10] and [11] to approximate the rate of evolution as $O(\log N / \log \log N)$. For other work in this direction see Rouzine, Brunet and Wilke [9], Brunet, Rouzine and Wilke [1], Desai and Fisher [2] and Park, Simon and Krug [8]. Using non-rigorous arguments, these authors get estimates of $O(\log N)$, $O(\log N / \log \log N)$, and $O(\log N / (\log \log N)^2)$, where the differences depend on the details of the models that they analyze. For more motivation and details concerning this model, please see the introduction in [11].

Motivated by applications to cancer development, Durrett and Mayberry have established rigorous results for a similar model in [4]. They consider two models in which all mutations are beneficial and the mutation rate tends to 0 as the population size tends to infinity. In one of their models the population size is fixed and in the other it is exponentially increasing. For the model with the fixed population size they show that the rate at which the average fitness is expected to increase is $O(\log N)$. By considering the expected number of individuals that have fitness k at time t , they establish rigorously that the density of the particles in their model will act as a traveling wave in time.

What we wish to establish in this paper is an upper bound for the mean rate of adaptation. We do this by bounding the rate at which the maximum fitness increases. Throughout the paper we let $X_t^+ = \max\{X_t^i : 1 \leq i \leq N\}$ be the maximum fitness of any cell at time t and $X_t^- = \min\{X_t^i : 1 \leq i \leq N\}$ be the minimum fitness of any cell at time t . Define the width of the process to be $W_t = X_t^+ - X_t^-$ and define $D_t = X_t^+ - X_0^+$ be the distance the front of the process has traveled by time t . Our main result is given in the following theorem.

Theorem 1. *Let $X_0^i = 0$ for $1 \leq i \leq N$. There exist positive constants a and b such that for N large enough*

$$\frac{1}{t} E[D_t] \leq (a + b\gamma) \frac{\log N}{\log \log N}$$

for all $t \geq 1$.

A difference between the result in [11] and our result is that in [11] the initial state of the process is randomly chosen according to the stationary distribution π while we make the assumption that all of the cells initially have fitness 0. With the initial condition $X_0^i = 0$ for $1 \leq i \leq N$, we let $\{\mathcal{F}_t\}_{t \geq 0}$ be the natural filtration associated with X .

Through the use of branching processes we establish a bound on D_t that depends on the width. If the width is large then our bound is also large. To establish our result we will need to show the width is not large too long or too often. This motivates us to make the following

definitions:

$$\begin{aligned}
t_1 &= 0 \\
s_n &= \inf\{t \geq t_n : W_t \geq 2 \log N\} \text{ for } n \geq 1 \\
t_n &= \inf\{t \geq s_{n-1} : W_t < \log N\} \text{ for } n \geq 2 \\
Y_i &= \sup_{s_i \leq t \leq t_{i+1}} D_t - D_{s_i} \text{ for } i \geq 1 \\
N_t &= \max\{i : s_i \leq t\} \text{ for } t \geq 0
\end{aligned}$$

We will make use of the strong Markov property of X at the times s_n and t_n for $n \geq 1$. For this reason, many of the statements we prove below will include conditions for which $W_0 > 0$ even though according to the conditions of Theorem 1 we have $W_0 = 0$. It is easy to see that the sequences $\{s_n\}_{n=1}^\infty$ and $\{t_n\}_{n=1}^\infty$ will be infinite with probability 1. We establish a small upper bound for D_t on the time intervals $[t_n, s_n)$ and show the complimentary time intervals $[s_n, t_{n+1})$ are infrequent enough that the fast rate of increase of D_t during these times will not significantly change the mean rate.

2 Coupling with Branching Processes

We begin this section by giving an equivalent description of the process X in which the Poisson processes that determine the dynamics of X are explicit. Fix the population size N . Then X is a stochastic process with state space \mathbb{Z}^n . The Poisson processes that determine the dynamics of X are as follows:

- There are N Poisson processes $\mathcal{P}^{i\uparrow}$, $1 \leq i \leq N$, on $[0, \infty)$ of rate $q\mu$. If $\mathcal{P}^{i\uparrow}$ gets a mark at t then the i^{th} coordinate of X increases by 1 at time t .
- There are N Poisson processes $\mathcal{P}^{i\downarrow}$, $1 \leq i \leq N$, on $[0, \infty)$ of rate $(1 - q)\mu$. If $\mathcal{P}^{i\downarrow}$ gets a mark at t then the i^{th} coordinate of X decreases by 1 at time t .
- For each ordered pair of coordinates (i, j) with $i \neq j$ there is a Poisson process on $[0, \infty)$, $\mathcal{P}^{i,j}$, of rate $1/N$. If $\mathcal{P}^{i,j}$ gets a mark at t then the j^{th} coordinate changes its value to agree with the i^{th} coordinate at time t .
- For each ordered pair of coordinates (i, j) with $i \neq j$ there is a Poisson processes on $[0, \infty) \times [0, \infty)$, $\mathcal{P}^{i,j\uparrow}$, which has intensity $\frac{\gamma}{N}\lambda$ where λ is Lebesgue measure on \mathbb{R}^2 . If $\mathcal{P}^{i,j\uparrow}$ gets a mark in $\{t\} \times [0, X_{t-}^i - X_{t-}^j]$ then the j^{th} coordinate changes its value to agree with the i^{th} coordinate at time t .

The following proofs will involve coupling X with various branching processes. To avoid confusion, we will refer to individuals in X as cells and we will refer to individuals in the branching processes as particles. While the cells in X each have an integer value that we refer to as the fitness of the cell, the particles in a branching process will each be given an integer value that we refer to as the type of the particle. Let $Z^C = \{Z_t^C\}_{t \geq 0}$ be a multi-type Yule process in which there are initially N particles of type 0. Particles increase from type i to type $i + 1$ at rate μ and

branch at rate C . When a particle of type i branches the new particle is also type i . Let M_t^C be the maximum type of any particle at time t .

Before proving a lemma about the branching process Z^C , we first prove a lemma which follows from Taylor's Remainder Theorem that we will make use of several times.

Lemma 2. *Let $x \geq 0$. The tail of the exponential series satisfies*

$$\sum_{i=k}^{\infty} \frac{x^i}{i!} \leq \frac{x^k e^x}{k!}.$$

Proof. By Taylor's Remainder Theorem we know that there exists a $\xi \in [0, x]$ such that

$$e^x = \sum_{i=1}^{k-1} \frac{x^i}{i!} + \frac{x^k e^\xi}{k!}.$$

Using the series expansion of e^x we have

$$\sum_{i=k}^{\infty} \frac{x^i}{i!} = \frac{x^k e^\xi}{k!} \leq \frac{x^k e^x}{k!}.$$

□

Lemma 3. *For any population size N , time $t \geq 0$, and natural number l ,*

$$P(M_t^C \geq l) \leq \frac{N(t\mu)^l e^{Ct}}{l!}.$$

Proof. Consider a Yule process Z which is the same as Z^C except there is only one particle at time 0. It is well known that the number of particles in Z_t has mean e^{Ct} . Let M'_t be the maximum type of any particle at time t . When there are k particles in the population, we let B_1, \dots, B_k denote the types of the particles, where the numbering is independent of the mutations. For any $l \geq 0$,

$$\begin{aligned} P(M'_t \geq l) &= \sum_{k=1}^{\infty} P(M'_t \geq l | Z_t = k) P(Z_t = k) \\ &= \sum_{k=1}^{\infty} P(\{B_1 \geq l\} \cup \dots \cup \{B_k \geq l\} | Z_t = k) P(Z_t = k) \\ &\leq \sum_{k=1}^{\infty} k P(B_1 \geq l) P(Z_t = k) \\ &= E[Z_t] P(B_1 \geq l) \\ &= e^{Ct} \sum_{i=l}^{\infty} \frac{(t\mu)^i}{i!} e^{-\mu t}. \end{aligned}$$

By Lemma 2 it follows that

$$P(M'_t \geq l) \leq \frac{(t\mu)^l e^{Ct}}{l!}.$$

Now consider Z^C . At time 0 label the particles $1, 2, \dots, N$ and let $M'_{i,t}$ be the maximum type of any particle among the progeny of particle i at time t . Then

$$\begin{aligned} P(M_t^C \geq l) &= P(\{M'_{1,t} \geq l\} \cup \dots \cup \{M'_{N,t} \geq l\}) \\ &\leq NP(M'_{1,t} \geq l) \\ &\leq \frac{N(t\mu)^l e^{Ct}}{l!}. \end{aligned}$$

□

The next proposition will give a lower bound on the fitness of any cell up to time t given that we know the least fitness at time 0 is X_0^- . We do this by establishing an upper bound on the amount that any cell will decrease in fitness. Let $S_t = \sup_{0 \leq s \leq t} (X_0^- - X_s^-)$.

Proposition 4. *For any population size N , time $t \geq 0$, and natural number l ,*

$$P(S_t \geq l) \leq \frac{N(t\mu)^l e^t}{l!}.$$

Proof. By Lemma 3 we have

$$P(M_t^1 \geq l) \leq \frac{N(t\mu)^l e^t}{l!}$$

for any population size N , time $t \geq 0$ and natural number l . To complete the proof we establish a coupling between X and Z^1 such that for any population size N and time $t \geq 0$ we have $M_t^1 \geq S_t$. At all times every cell in X will be paired with one particle in Z^1 . The coupling is as follows:

- We initially have a one-to-one pairing of each cell i in X_0 with each particle i in Z_0^1 .
- The particle in Z^1 that is paired with cell i will increase in type by 1 only when cell i gets a mutation.
- For each cell i in X , at rate $(N-1)/N$ cells $j \neq i$ are replaced by cell i due to resampling events. If cell i replaces cell j due to resampling, then the particle labeled i in Z^1 branches. If particle i has a higher type than particle j then the new particle is paired with cell j . The particle that was paired with cell j before the branching event is no longer paired with any cell in X . If particle i has a lower type than particle j then the particle that was paired with cell j remains paired with cell j and the new particle is not paired with any cell in X .
- The particle paired with cell i in Z^1 branches at rate $1/N$ and these branching events are independent of any of the events in X . When the particle paired with cell i branches due to these events the new particle is not paired with any cell in X .
- Any particles in Z^1 that are not paired with a cell in X branch and acquire mutations independently of X . The selection events in X are independent of any events in Z^1 .

Let R_s^i be the type of the particle in Z_s^1 that is paired with cell i . To bound S_t we consider $S_s^i = \sup_{0 \leq r \leq s} (X_0^- - X_r^i)$ for $1 \leq i \leq N$. To show $M_t^1 \geq S_t$ it is enough to show $R_t^i \geq S_t^i$ for all i . Initially $S_0^i \leq R_0^i = 0$ for all i . Note that both $s \mapsto S_s^i$ and $s \mapsto R_s^i$ are increasing functions and that increases in these functions correspond to decreases in X_s^i .

When cell i gets a mutation, R^i increases by 1. However, if cell i gets a mutation at time s then S^i will only increase by 1 if $S_{s-}^i = X_0^- - X_{s-}^i$ and the mutation is deleterious. Therefore, if cell i gets a mutation at time s and $S_{s-}^i \leq R_{s-}^i$ then

$$S_s^i \leq S_{s-}^i + 1 \leq R_{s-}^i + 1 = R_s^i.$$

Suppose cell j is replaced by cell i due to a resampling event at time s and that both $S_{s-}^j \leq R_{s-}^j$ and $S_{s-}^i \leq R_{s-}^i$ hold. With probability 1 we have $S_s^j = S_{s-}^j$ and $R_s^j = R_{s-}^j$. If $X_0^- - X_s^i \leq S_{s-}^j$ then $S_s^j = S_s^i$. From this it follows that $S_s^j \leq R_s^i$. If $X_0^- - X_s^i > S_{s-}^j$ then $S_s^j = X_0^- - X_s^i \leq S_s^i \leq R_s^i$. If $R_s^i \geq R_{s-}^j$, then by the definition of the coupling, $R_s^j = R_{s-}^j$. If $R_s^i < R_{s-}^j$, then by definition of the coupling, $R_s^j = R_{s-}^j$. Therefore, $R_s^j \geq R_s^i$ which gives us $S_s^j \leq R_s^j$.

Selection events will never increase S^i and since S^i and R^i are increasing in time a selection event at time s will preserve the inequality $S_s^i \leq R_s^i$. This shows that any event that occurs at time s which may change the fitness of a cell i in X will preserve the inequality $S_s^i \leq R_s^i$. Since the result holds for each individual i we have $S_t \leq M_t^1$. \square

We now wish to bound the distance the front of the wave moves as a function of the initial width. To do so we define a new branching process $Z^{k,\uparrow}$ which has the following dynamics:

- Initially there are N particles of type k in $Z_0^{k,\uparrow}$.
- Each particle changes from type n to $n + 1$ at rate μ .
- A particle of type n branches at rate $\gamma n + 1$ and upon branching the new particle is also type n .

Let $\bar{M}_t^{k,\uparrow}$ be the maximum type of any particle in $Z_t^{k,\uparrow}$ and let $M_t^{k,\uparrow} = \bar{M}_t^{k,\uparrow} - k$, so that $M_0^{k,\uparrow} = 0$.

Lemma 5. *For any time $t \geq 0$ and any integers $k \geq 0$ and $l \geq 0$ we have*

$$P(M_t^{k,\uparrow} > l) \leq \frac{N(t\mu)^l e^{(\gamma(k+l)+1)t}}{l!}.$$

Proof. While all of the particles in $Z_t^{k,\uparrow}$ have type less than $k + l$ they branch at a rate which is less than or equal to $\gamma(k + l) + 1$. Because of this, $P(M_t^{k,\uparrow} > l) \leq P(M_t^{\gamma(k+l)+1} > l)$. By Lemma 3 we have

$$P(M_t^{\gamma(k+l)+1} > l) \leq \frac{N(t\mu)^l e^{(\gamma(k+l)+1)t}}{l!}.$$

\square

Proposition 6. *For any fixed time $t \geq 0$ and any integer $l \geq 0$ we have*

$$P(\sup_{0 \leq s \leq t} D_s > l) \leq \frac{2N(t\mu)^l e^{(\gamma(W_0+2l)+\mu+1)t}}{(l-1)!}.$$

Proof. We first establish a coupling between X and $Z^{W_0+k,\uparrow}$ for each integer $k \geq 0$. For $k \geq 1$ define $T^k = \inf\{t : S_t > k\}$. Every cell in X will be paired with one particle in $Z^{W_0+k,\uparrow}$ until time T^k . We couple $Z^{W_0+k,\uparrow}$ with X for all times $t \in [0, T^k)$ as follows:

- We initially have a one-to-one pairing of each cell i in X_0 with each particle i in $Z_0^{W_0+k,\uparrow}$.
- The particle in $Z^{W_0+k,\uparrow}$ that is paired with cell i will increase in type by 1 only when cell i gets a mutation.
- For each cell i in X , at rate $(N-1)/N$ cells $j \neq i$ are replaced by cell i due to resampling events. If cell i replaces cell j due to resampling, then the particle labeled i in $Z^{W_0+k,\uparrow}$ branches. If particle i has a higher type than particle j then the new particle is paired with cell j . The particle that was paired with cell j before the branching event is no longer paired with any cell in X . If particle i has a lower type than particle j then the particle that was paired with cell j remains paired with cell j and the new particle is not paired with any cell in X .
- Additionally, the particle paired with cell i in $Z^{W_0+k,\uparrow}$ branches at rate $1/N$ and these branching events are independent of any of the events in X . When the particle paired with cell i branches due to these events the new particle is not paired with any cell in X .
- In X there is a time dependent rate γU_s^i at which cells $j \neq i$ are replaced by cell i due to selection events. If cell j is replaced by cell i in X due to a selection event then the particle labeled i in $Z_t^{W_0+k,\uparrow}$ splits. If particle i has a higher type than particle j then the new particle is paired with cell j . The particle that was paired with cell j before the branching event is no longer paired with any cell in X . If particle i has a lower type than particle j then the particle that was paired with cell j remains paired with cell j . The new particle is not paired with any cell in X .
- Additionally, a particle labeled i in $Z^{W_0+k,\uparrow}$ splits at a time dependent rate $\gamma(R_t^{i,k} - U_t^i)$ where $R_t^{i,k}$ is the type of particle i . These branching events are independent of any of the events in X . When such a branching event occurs, the new particle is not paired with any cell in X .
- Any particles in $Z^{W_0+k,\uparrow}$ that are not paired with a cell in X branch and acquire mutations independently of X .

Fix $k \geq 1$. For the above coupling between X and $Z^{W_0+k,\uparrow}$ to be well defined until time T^k , we need $R_t^{i,k} - U_t^i \geq 0$ for all $i \in \{1, \dots, N\}$ and for all times $t \in [0, T^k]$. Let

$$\bar{T}^{k,i} = \inf\{t : R_t^{i,k} - U_t^i < 0\}.$$

The coupling between X and $Z^{W_0+k,\uparrow}$ is well-defined until time $\bar{T}^k = \min\{\bar{T}^{k,i} : 1 \leq i \leq N\}$. We will show that $T^k \leq \bar{T}^k$.

Let $\bar{S}_t^i = \sup_{0 \leq s \leq t} (X_s^i - X_0^+)$ and let $\bar{R}_t^{i,k} = R_0^{i,k} - W_0 - k$. Initially $\bar{S}_0^i \leq \bar{R}_0^{i,k} = 0$ for all i . Note that both $t \mapsto \bar{S}_t^i$ and $t \mapsto R_t^{i,k}$ are increasing functions, from which it follows that $t \mapsto \bar{R}_t^{i,k}$ is also an increasing function.

When cell i gets a mutation, $\bar{R}_t^{i,k}$ increases by 1. However, if cell i gets a mutation at time s then \bar{S}_s^i will only increase by 1 if $\bar{S}_{s-}^i = X_{s-}^i - X_0^+$ and the mutation is beneficial. Therefore, if cell i gets a mutation at time s and $\bar{S}_{s-}^i \leq \bar{R}_{s-}^{i,k}$ then

$$\bar{S}_s^i \leq \bar{S}_{s-}^i + 1 \leq \bar{R}_{s-}^{i,k} + 1 = \bar{R}_s^{i,k}.$$

Suppose cell j is replaced by cell i due to a resampling or selection event at time s and that both $\bar{S}_{s-}^j \leq \bar{R}_{s-}^{j,k}$ and $\bar{S}_s^i = \bar{S}_{s-}^i \leq \bar{R}_{s-}^{i,k} = \bar{R}_s^{i,k}$ hold. If $X_s^i - X_0^+ \leq \bar{S}_{s-}^j$ then $\bar{S}_{s-}^j = \bar{S}_s^j$. It follows that $\bar{S}_s^j \leq \bar{R}_s^{j,k}$. If $X_s^i - X_0^+ > \bar{S}_{s-}^j$ then $\bar{S}_s^j = X_0^- - X_s^i \leq \bar{S}_s^i \leq \bar{R}_s^{i,k}$. If $\bar{R}_s^{i,k} \geq \bar{R}_{s-}^{j,k}$, then by the definition of the coupling, $\bar{R}_s^{i,k} = \bar{R}_s^{j,k}$. If $\bar{R}_s^{i,k} < \bar{R}_{s-}^{j,k}$, then by definition of the coupling, $\bar{R}_s^{j,k} = \bar{R}_s^{i,k}$. Therefore, $\bar{R}_s^{j,k} \geq \bar{R}_s^{i,k}$ which gives us $\bar{S}_s^j \leq \bar{R}_s^{j,k}$.

For any time $s < T^k$ we have $R_s^{i,k} \geq \bar{S}_s^i + W_0 + k \geq X_s^i - X_0^+ + W_0 + k = X_s^i - X_0^- + k$. If there were N cells with fitness $X_0^- - k$ at time $s \in [0, \bar{T}^{k,i})$ then the rate at which cell i replaces these N cells due to selection is $\gamma(X_s^i - X_0^0 + k)$. However, for any time $s < T^k$ there are fewer than N cells being replaced by cell i due to selection and they will all have fitnesses at least as large as $X_0^- - k$. This gives us a bound on the rate at which resampling events occur on cell i before time T^k , namely $U_s^i \leq X_s^i - X_0^- + k \leq R_s^{i,k}$ for all $s \in [0, T^k)$. This shows that $T^k \leq \bar{T}^{k,i}$ for all i . Hence, $T^k \leq \bar{T}^k$ and the coupling is well-defined until time T^k .

We have shown that any event that occurs at time $s \in [0, T^k)$ which may change the fitness of a cell i in X will preserve the inequality $\bar{S}_s^i \leq \bar{R}_s^{i,k}$. Since the result holds for each individual i , for any $s \in [0, T^k)$ we have

$$\sup_{0 \leq r \leq s} D_r = \sup_{1 \leq i \leq N} \bar{S}_s^i \leq \sup_{1 \leq i \leq N} \bar{R}_s^{i,k} \leq M_s^{W_0+k,\uparrow}.$$

Note that if $\sup_{0 \leq s \leq t} (X_0^- - X_s^-) \leq k$ then $t < T^k$. Therefore, on the event

$$\{ \sup_{0 \leq s \leq t} (X_0^- - X_s^-) \leq k \}$$

we have $M_t^{W_0+k,\uparrow} \geq \sup_{0 \leq s \leq t} D_s$. This allows us to do the following computation:

$$\begin{aligned} P(\sup_{0 \leq s \leq t} D_s > l) &= \sum_{i=0}^{\infty} P(\{\sup_{0 \leq s \leq t} D_s > l\} \cap \{\sup_{0 \leq s \leq t} (X_0^- - X_s^-) = i\}) \\ &\leq \sum_{i=0}^{\infty} P(\{M_t^{W_0+i,\uparrow} > l\} \cap \{\sup_{0 \leq s \leq t} (X_0^- - X_s^-) = i\}) \\ &\leq \sum_{i=0}^{\infty} P(\{M_t^{W_0+i,\uparrow} > l\} \cap \{\sup_{0 \leq s \leq t} (X_0^- - X_s^-) \geq i\}) \\ &\leq \sum_{i=0}^{\infty} P(M_t^{W_0+i,\uparrow} > l) \wedge P(\sup_{0 \leq s \leq t} (X_0^- - X_s^-) \geq i) \\ &\leq \sum_{i=0}^{\infty} P(M_t^{W_0+i,\uparrow} > l) \wedge \left(\frac{N(t\mu)^i e^t}{i!} \right) \text{ by Proposition 4} \\ &\leq \sum_{i=0}^{\infty} \left(\frac{N(t\mu)^l e^{(\gamma(W_0+i+l)+1)t}}{l!} \right) \wedge \left(\frac{N(t\mu)^i e^t}{i!} \right) \text{ by Lemma 5} \\ &\leq \frac{N(t\mu)^l e^{(\gamma(W_0+l)+1)t}}{l!} \sum_{i=0}^{l-1} e^{i\gamma t} + N e^t \sum_{i=l}^{\infty} \frac{(t\mu)^i}{i!} \end{aligned}$$

$$\begin{aligned}
&\leq \frac{N(t\mu)^l e^{(\gamma(W_0+l)+1)t}}{l!} \cdot l e^{l\gamma t} + N e^t \sum_{i=l}^{\infty} \frac{(t\mu)^i}{i!} \\
&\leq \frac{N(t\mu)^l e^{(\gamma(W_0+2l)+1)t}}{(l-1)!} + \frac{N(t\mu)^l e^{(\mu+1)t}}{l!} \text{ by Lemma 2} \\
&\leq \frac{2N(t\mu)^l e^{(\gamma(W_0+2l)+\mu+1)t}}{(l-1)!}.
\end{aligned} \tag{1}$$

□

We now extend the bound we got on the least fit individuals in Proposition 4 to a slightly stronger result. Let $x \in \mathbb{Z}$ and let $\bar{S}_t^x \subset \{1, 2, \dots, N\}$ correspond to a collection of cells at time t which is determined by the following dynamics:

- Initially \bar{S}_0^x consists of all cells whose fitness lies in the interval (x, ∞) .
- If a resampling or selection event occurs at time t and a cell not in \bar{S}_{t-}^x is replaced by a cell in \bar{S}_{t-}^x then it is added to \bar{S}_t^x .
- If a beneficial mutation occurs at time t on a cell not in \bar{S}_{t-}^x that causes its fitness to increase from x to $x+1$ it is added to \bar{S}_t^x .
- If a resampling event occurs at time t to a cell in \bar{S}_{t-}^x and it is replaced by a cell not in \bar{S}_{t-}^x then it is removed from \bar{S}_t^x .

Mutation and selection events do not cause cells to be lost from \bar{S}^x . We now prove the following corollary to Proposition 6.

Corollary 7. *Let A_t^l be the event that a cell in \bar{S}_s^x has fitness in $(-\infty, x-l]$ for some time $s \in [0, t]$. For any time $t \geq 0$ and any integer l ,*

$$P(A_t^l) \leq \frac{2N(t\mu)^l e^{(\gamma(W_0+2l)+\mu+1)t}}{(l-1)!}.$$

Note that we cannot use the bound found in Proposition 4 because cells not in \bar{S}_t^x may move to \bar{S}_t^x due to selection events. Proposition 4 does not say anything about what happens if the population of cells with lower fitness in \bar{S}^x is increasing rapidly due to selection events with cells of even lower fitness outside of \bar{S}^x .

Proof of Corollary 7. For $k \geq 1$ let X be coupled with $Z^{W_0+k,\uparrow}$ as in the proof of Proposition 6. Let T^k , $R_t^{i,k}$ and $\bar{R}_t^{i,k}$ be defined as they were in the proof of Proposition 6. Define

$$\bar{T}_s^i = \{r \in [0, s] : i \in S_r^x\}$$

and let

$$S_s^i = \begin{cases} \sup_{r \in \bar{T}_s^i} (x - X_r^i) & \text{if } \bar{T}_s^i \neq \emptyset \\ -\infty & \text{if } \bar{T}_s^i = \emptyset \end{cases}.$$

The goal is to show that for all $s \in [0, T^k)$ we have

$$\sup_{1 \leq i \leq N} S_s^i \leq \sup_{1 \leq i \leq N} \bar{R}_s^{i,k} \leq M_s^{W_0+k,\uparrow}.$$

Note that we can only consider the coupling of X with $Z^{W_0+k,\uparrow}$ until time T^k because after this time the coupling is not well-defined.

Initially all of the cells in \bar{S}_0^x have fitness in (x, ∞) . Therefore, if $i \in \bar{S}_0^x$ then $S_0^i \leq 0 = \bar{R}_0^{i,k}$. If $i \notin \bar{S}_0^x$ then $S_0^i = -\infty < \bar{R}_0^{i,k}$.

Suppose cell i gets a mutation at time s and that for any time $s' \in [0, s-)$ we have $S_{s'}^i \leq \bar{R}_{s'}^{i,k}$. Then $\bar{R}^{i,k}$ increases by 1. If $i \in \bar{S}_{s-}^x$ then S_s^i will only increase by 1 if $S_{s-}^i = x - X_s^i$ and the mutation is deleterious. If $i \notin \bar{S}_{s-}^x$ and the mutation does not cause the fitness of cell i to change from x to $x+1$ then $S_s^i = S_{s-}^i$. If $i \notin \bar{S}_{s-}^x$ and the mutation does cause the fitness of cell i to change from x to $x+1$ then $S_s^i = S_{s-}^i \vee 0$. In any of these three cases, $S_s^i \leq \bar{R}_s^{i,k}$.

Suppose cell j is replaced by cell i due to a resampling or selection event at time s and that $S_{s-}^j \leq \bar{R}_{s-}^{j,k}$ and $S_{s-}^i \leq \bar{R}_{s-}^{i,k}$. If $i \notin \bar{S}_{s-}^x$ then $S_{s-}^j = S_s^j \leq \bar{R}_{s-}^{j,k}$. Suppose $i \in \bar{S}_{s-}^x$. If $x - X_s^i \leq S_{s-}^j$ then $S_{s-}^j = S_s^j$. From this it follows that $S_s^j \leq \bar{R}_s^j$. If $x - X_s^i > S_{s-}^j$ then $S_s^j = x - X_s^i \leq S_s^i \leq \bar{R}_s^i$. If $\bar{R}_s^i \geq \bar{R}_{s-}^j$, then by the definition of the coupling, $\bar{R}_s^j = \bar{R}_s^i$. If $\bar{R}_s^i < \bar{R}_{s-}^j$, then by definition of the coupling, $\bar{R}_s^j = \bar{R}_{s-}^j$. Therefore, $\bar{R}_s^j \geq \bar{R}_s^i$ which gives us $S_s^j \leq \bar{R}_s^j$.

Note that if $\sup_{0 \leq s \leq t} (X_0^- - X_s^-) \leq k$ then $t < T^k$. Therefore, on the event

$$\left\{ \sup_{0 \leq s \leq t} (X_0^- - X_s^-) \leq k \right\}$$

we have $M_t^{W_0+k,\uparrow} \geq \sup_{1 \leq i \leq N} S_s^i$. This allows us to do the following computation:

$$\begin{aligned} P\left(\sup_{0 \leq s \leq t} \sup_{1 \leq i \leq N} S_s^i > l\right) &= \sum_{i=0}^{\infty} P\left(\left\{ \sup_{0 \leq s \leq t} \sup_{1 \leq i \leq N} S_s^i > l \right\} \cap \left\{ \sup_{0 \leq s \leq t} (X_0^- - X_s^-) = i \right\}\right) \\ &\leq \sum_{i=0}^{\infty} P\left(\left\{ M_t^{W_0+i,\uparrow} > l \right\} \cap \left\{ \sup_{0 \leq s \leq t} (X_0^- - X_s^-) = i \right\}\right). \end{aligned}$$

This is the same bound as equation (1) in the proof of Proposition 6. Therefore, we have established the same bound. \square

3 Bounding the rate when the width is small

We define a process X' which we will couple with X when the width is small. To define X' we first define an i.i.d. sequence of continuous-time stochastic processes $\{\mathcal{Z}^n\}_{n=0}^{\infty}$ where the distribution of each \mathcal{Z}^n is the same as the distribution of $Z^{2 \log N, \uparrow}$. Let $\bar{\mathcal{M}}_t^n$ be the maximum type of any particle in \mathcal{Z}_t^n and let $\mathcal{M}_t^n = \bar{\mathcal{M}}_t^n - 2 \log N$ so that $\mathcal{M}_0^n = 0$ for all n . Define

$$X'_t = \begin{cases} X_0^+ + \mathcal{M}_t^0 & \text{if } t \in [0, 1] \\ X'_{\lfloor t \rfloor} + \mathcal{M}_{t-\lfloor t \rfloor}^i & \text{if } t \in (i, i+1] \text{ for any integer } i \geq 1 \end{cases}$$

and $D'_t = X'_t - X_0^+$. The idea is that D'_t is the maximum type of any particle in a branching process X' that has the same distribution as $Z^{2 \log N, \uparrow}$ except that at each integral time unit we restart the branching process so that there are once again N particles. At each integral time unit t the N particles initially have type D'_t which is the maximum type achieved by any particle in X'_t up to time t .

Proposition 8. Let $\alpha > 1 + 2\gamma$. For N large enough we have

$$\sup_{t \in [1, \infty)} \frac{1}{t} E[D'_t] \leq 2\alpha \bar{k}.$$

Proof. By definition D'_1 has the same distribution as $M_1^{2 \log N, \uparrow}$ so by Lemma 5 we have

$$P(D'_1 > l) \leq \frac{N^{1+2\gamma} \mu^l e^{\gamma l + 1}}{l!}.$$

Then

$$\frac{E[D'_1]}{\bar{k}} = \frac{1}{\bar{k}} \sum_{l=0}^{\infty} P(D'_1 > l) \leq \frac{1}{\bar{k}} \left[\alpha \bar{k} + \sum_{l=\lceil \alpha \bar{k} \rceil}^{\infty} \frac{N^{1+2\gamma} \mu^l e^{\gamma l + 1}}{l!} \right].$$

By Lemma 2 we have

$$\frac{E[D'_1]}{\bar{k}} \leq \alpha + \frac{N^{1+2\gamma} e}{\bar{k}} \sum_{l=\lceil \alpha \bar{k} \rceil}^{\infty} \frac{\mu^l e^{\gamma l}}{l!} \leq \alpha + \frac{N^{1+2\gamma} \mu^{\alpha \bar{k}} e^{\alpha \gamma \bar{k} + \mu e^\gamma}}{\bar{k} \lceil \alpha \bar{k} \rceil!}.$$

Note that for any $k \geq 2$ both $D'_k - D'_{k-1}$ and D'_1 have the same distribution, namely that of $M_1^{2 \log N, \uparrow}$. Choose $t \in [k, k+1)$ for some $k \geq 1$. Because D'_t is increasing in t , we have

$$\frac{1}{t} D'_t \leq \frac{1}{k} (D'_{k+1} - D'_k + D'_k - \cdots + D'_2 - D'_1 + D'_1).$$

Therefore,

$$\frac{1}{t} E[D'_t] \leq \frac{k+1}{k} E[D'_1] < 2E[D'_1].$$

Dividing each side by \bar{k} and using the bound on $E[D'_1]$ gives us

$$\frac{1}{\bar{k}t} E[D'_t] \leq 2\alpha + \frac{2N^{1+2\gamma} \mu^{\alpha \bar{k}} e^{\alpha \gamma \bar{k} + \mu e^\gamma}}{\bar{k} \lceil \alpha \bar{k} \rceil!}$$

By Stirling's formula we have

$$\frac{2N^{1+2\gamma} \mu^{\alpha \bar{k}} e^{\alpha \gamma \bar{k} + \mu e^\gamma}}{\bar{k} \lceil \alpha \bar{k} \rceil!} \sim \frac{N^{1+2\gamma} \mu^{\alpha \bar{k}} e^{\alpha \gamma \bar{k} + \lceil \alpha \bar{k} \rceil + \mu e^\gamma}}{\sqrt{2\pi \bar{k}} \lceil \alpha \bar{k} \rceil^{\lceil \alpha \bar{k} \rceil + 1/2}} \rightarrow 0 \text{ as } N \rightarrow \infty,$$

as long as $\alpha > 1 + 2\gamma$. □

Proposition 9. Let $X_0^i = 0$ for $1 \leq i \leq N$. Then

$$D_t \leq D'_t + \sum_{i=1}^{N_t} Y_i$$

for all times $t \geq 0$.

Proof. We now couple X with X' by coupling X with the sequence of processes $\{\mathcal{Z}^m\}_{m=0}^\infty$. Let

$$I_m = (m, m+1] \cap \bigcup_{n=1}^{\infty} [t_n, s_n) \text{ and } J_m = (0, 1] \cap \bigcup_{n=1}^{\infty} [t_n - m, s_n - m).$$

For any $m \geq 0$ we couple X and \mathcal{Z}^m as follows:

- The particles in \mathcal{Z}_0^m are labeled $1, 2, \dots, N$.
- For any time in I_m^C the process X behaves independently of \mathcal{Z}^m . For any time in J_m^C the process \mathcal{Z}^m behaves independently of the process X . During the time J_m^C , if a particle labeled i in \mathcal{Z}^m branches, the particle remains labeled i and the new particle is unlabeled.
- The particle in \mathcal{Z}^m that is paired with cell i will increase in type by 1 at time $t \in J_m$ only when cell i gets a mutation at time $t + m \in I_m$.
- For each cell i in X , at rate $(N-1)/N$ cells $j \neq i$ are replaced by cell i due to resampling events. If cell i replaces cell j due to resampling at time $t \in I_m$ then the particle labeled i in \mathcal{Z}^m branches at time $t - m \in J_m$. If particle i has a higher type than particle j then the new particle is paired with cell j . The particle that was paired with cell j before the branching event is no longer paired with any cell in X . If particle i has a lower type than particle j then the particle that was paired with cell j remains paired with cell j and the new particle is not paired with any cell in X .
- The particle paired with cell i in \mathcal{Z}^m branches at rate $1/N$ for all times $t \in J_m$ and these branching events are independent of any of the events in X . When the particle paired with cell i branches due to these events the new particle is not paired with any cell in X .
- In X there is a time dependent rate γU_s^i at which cells $j \neq i$ are replaced by cell i due to selection events. If cell j is replaced by cell i in X due to a selection event at time $t \in I_m$ then the particle labeled i in \mathcal{Z}^m splits at time $t - m \in J_m$. If particle i has a higher type than particle j then the new particle is paired with cell j . The particle that was paired with cell j before the branching event is no longer paired with any cell in X . If particle i has a lower type than particle j then the particle that was paired with cell j remains paired with cell j . The new particle is not paired with any cell in X .
- A particle labeled i in \mathcal{Z}^m splits at a time dependent rate $\gamma(R_t^{i,k} - U_t^i)$ for all times $t \in J_m$ where $R_t^{i,k}$ is the type of particle i . These branching events are independent of any of the events in X . When such a branching event occurs, the new particle is not paired with any cell in X .
- Any particles in \mathcal{Z}^m that are not paired with a cell in X branch and acquire mutations independently of X .

Observe the following bound for D_t :

$$D_t \leq \sum_{i=1}^{N_t-1} (D_{t_{i+1}} - D_{s_i}) + \sum_{i=1}^{N_t} (D_{s_i} - D_{t_i}) + \sup_{s_{N_t} \leq s \leq t_{N_t+1}} (D_s - D_{s_{N_t}}) + \sup_{t_{N_t+1} \leq s \leq t} (D_s - D_{t_{N_t+1}}),$$

where we consider the supremum over the empty set to be 0. By definition we have

$$\sum_{i=1}^{N_t-1} (D_{t_{i+1}} - D_{s_i}) + \sup_{s_{N_t} \leq s \leq t_{N_t+1}} (D_s - D_{s_{N_t}}) \leq \sum_{i=1}^{N_t} Y_i.$$

To finish the proof we will show

$$\sum_{i=1}^{N_t} \sup_{t_i \leq s \leq s_i} (D_s - D_{t_i}) + \sup_{t_{N_t+1} \leq s \leq t} (D_s - D_{t_{N_t+1}}) \leq D'_t.$$

To do this we define

$$M_t = \sum_{i=1}^{N_t} \sup_{t_i \leq s \leq s_i} (D_s - D_{t_i}) + \sup_{t_{N_t+1} \leq s \leq t} (D_s - D_{t_{N_t+1}})$$

for all times $t \geq 0$. Suppose $M_s \leq D'_s$ for all $s \in [0, t)$ and a mutation, resampling or selection event occurs in X at time t . If $t \in (s_i, t_{i+1})$ for some $i \geq 0$ then $M_{t-} = M_t$ because the process M does not change on these time intervals. It is possible that D'_t changes, but D'_t can only increase. Therefore, $D'_t \geq M_t$. If $t \in [t_i, s_i] \cap (m, m+1]$ for some $i \geq 0$ and $m \geq 0$ then at time t the processes X and X' are coupled. More precisely, X and Z^m are coupled and the coupling has the same dynamics as the coupling in Proposition 6 except the time shift. The same argument used in Proposition 6 shows that $D'_t \geq M_t$ whether the cell changed fitness due to mutation, resampling or selection. Since this inequality is preserved on any event that may change M_t , it is true for all times t . \square

4 Bounding $E[Y_i | \mathcal{F}_{s_i}]$

For this section we let $f(N) = \log \log \log N$.

The first goal of this section is to show that when the width is large enough the selection mechanism will cause the width to decrease quickly. We give a labeling to the cells that will help us in this regard. Define the following subsets of \mathbb{R} :

$$\begin{aligned} I_1 &= (-\infty, X_0^+ - \frac{3}{16}W_0] \\ I_2 &= (X_0^+ - \frac{3}{16}W_0, X_0^+ - \frac{2}{16}W_0] \\ I_3 &= (X_0^+ - \frac{2}{16}W_0, X_0^+ - \frac{1}{16}W_0] \\ I_4 &= (X_0^+ - \frac{1}{16}W_0, \infty) \end{aligned}$$

We will label each cell in X_0 with two labels. For the first labeling, we use \mathfrak{a} to label the cells in $I_1 \cup I_2$, we use \mathfrak{b} to label the cells in I_3 and we use \mathfrak{c} to label the cells in I_4 . For the second labeling we use \mathfrak{a}' to label the cells in I_1 , we use \mathfrak{b}' to label the cells in I_2 and we use \mathfrak{c}' to label the cells in $I_3 \cup I_4$.

Let \mathfrak{A}_t , \mathfrak{B}_t and \mathfrak{C}_t denote the number of cells labeled \mathfrak{a} , \mathfrak{b} and \mathfrak{c} at time t , respectively. Let \mathfrak{A}'_t , \mathfrak{B}'_t and \mathfrak{C}'_t denote the number of cells labeled \mathfrak{a}' , \mathfrak{b}' and \mathfrak{c}' at time t , respectively.

The cells change labels over time according to the following dynamics:

- Mutations: If the fitness of a cell labeled \mathfrak{a} increases so that it is in I_3 then the cell is relabeled \mathfrak{b} . If the fitness of a cell labeled \mathfrak{a}' increases so that it is in I_2 then the cell is relabeled \mathfrak{b}' . Likewise, if the fitness of a cell labeled \mathfrak{b} increases so that it is in I_4 then it is relabeled \mathfrak{c} and if the fitness of a cell labeled \mathfrak{b}' increases so that it is in I_3 then it is relabeled \mathfrak{c}' . Deleterious mutations do not cause cells to be relabeled.
- Resampling: Any resampling event in which cell i is replaced by cell j causes cell i to inherit the labels of cell j .
- Selection: If a cell labeled \mathfrak{a} is replaced due to a selection event it inherits the corresponding label of the cell that replaced it. If a cell labeled \mathfrak{a}' is replaced due to a selection event it inherits the corresponding label of the cell that replaced it. If a cell labeled \mathfrak{b} is replaced by a cell labeled \mathfrak{c} due to a selection event then the cell that was labeled \mathfrak{b} is relabeled \mathfrak{c} . If a cell labeled \mathfrak{b}' is replaced by a cell labeled \mathfrak{c}' due to a selection event then the cell that was labeled \mathfrak{b}' is relabeled \mathfrak{c}' . Any other selection events do not cause the labels of the cells to be changed.

Let A_1 be the event that there is a cell labeled \mathfrak{b} with fitness in $(-\infty, X_0^+ - \frac{5}{32}W_0)$ for some time $t \in [0, f(N)]$. Let A_2 be the event that there is a cell labeled \mathfrak{c} with fitness in $(-\infty, X_0^+ - \frac{3}{32}W_0)$ for some time $t \in [0, f(N)]$. Let A'_1 be the event that there is a cell labeled \mathfrak{b}' with fitness in $(-\infty, X_0^+ - \frac{7}{32}W_0)$ for some time $t \in [0, f(N)]$. Let A'_2 be the event that there is a cell labeled \mathfrak{c}' with fitness in $(-\infty, X_0^+ - \frac{5}{32}W_0)$ for some time $t \in [0, f(N)]$.

Lemma 10. *Consider any sequence of initial configurations X_0 such that $W_0 \geq \log N$ for all N . Then*

$$P(A_1 \cup A_2 \cup A'_1 \cup A'_2) \rightarrow 0 \text{ as } N \rightarrow \infty.$$

Proof. First we show the result for A_1 . We apply Corollary 7 with $x = X_0^+ - 2W_0/16$, $t = f(N)$ and $l = W_0/32$. Because $x = X_0^+ - 2W_0/16$ we have that \bar{S}_0^x consists of all the cells labeled \mathfrak{b} or \mathfrak{c} . Setting $t = f(N)$ and $l = W_0/32$ will make A_t^l the event that a cell labeled \mathfrak{b} or \mathfrak{c} has fitness less than $X_0^+ - \frac{5}{32}W_0$ by time $f(N)$. Note that according to the relabeling dynamics, cell i being labeled \mathfrak{b} or \mathfrak{c} is equivalent to $i \in \bar{S}^x$. Therefore, $A_1 \subset A_t^l$ and we get

$$P(A_1) \leq P(A_t^l) \leq \frac{2N(t\mu)^l e^{(\gamma(W_0+2l)+\mu+1)t}}{(l-1)!} \rightarrow 0 \text{ as } N \rightarrow \infty.$$

We can apply Corollary 7 with $x = X_0^+ - 1/16$, $t = f(N)$ and $l = W_0/32$ to get the same bound for $P(A_2)$. By choosing x , t and l in this way the event A_t^l is the event that a cell labeled \mathfrak{c} has fitness less than $X^+(0) - \frac{3}{32}W_0$ by time $f(N)$. This shows that $P(A_2)$ also tends to 0 as N tends to infinity.

Likewise, to show $P(A'_1)$ tends to 0 as N goes to infinity we can apply Corollary 7 with $x = X_0^+ - \frac{3}{16}W_0$, $t = f(N)$ and $l = W_0/32$, and to show $P(A'_2)$ tends to 0 as N goes to infinity we can apply Corollary 7 with $x = X_0^+ - \frac{2}{16}W_0$, $t = f(N)$ and $l = W_0/32$. \square

Lemma 11. *Consider any sequence of initial configurations X_0 such that $W_0 \geq \log N$ for all N . Let T be a stopping time whose definition may depend on N such that $\mathfrak{C}'_T \geq N/4$ for all N . Let $B_T = \inf\{t \geq T : X_t^- > X_0^+ - W_0/4\}$. Then*

$$P(B_T 1_{\{T < \frac{1}{2}f(N)\}} > \frac{1}{2}f(N)) \rightarrow 0 \text{ as } N \rightarrow \infty.$$

Proof. Let A'_3 be the event that $\mathfrak{C}'_t \geq N/5$ for all times $t \in [T, T + \frac{1}{2}f(N)]$. The only way for a cell labeled \mathfrak{c}' to change its label is for it to be replaced by a cell labeled \mathfrak{a}' or \mathfrak{b}' via a resampling event. The rate at which cells marked \mathfrak{c}' undergo resampling events with cells marked \mathfrak{a}' or \mathfrak{b}' at time t is

$$\frac{\mathfrak{C}'_t(N - \mathfrak{C}'_t)}{N} \leq \frac{N}{4}.$$

Let $\{U_n\}_{n=0}^\infty$ be a simple random walk with $U_0 = N/4 \leq \mathfrak{C}'_T$. Let $T \leq t_1 < t_2 < \dots$ be the times at which cells labeled \mathfrak{c}' are involved in resampling events with cells that are not labeled \mathfrak{c}' after time T . We couple $\{U_n\}_{n=0}^\infty$ with X so that if at time t_n a cell is labeled \mathfrak{c}' due to a resampling event then $U_n = U_{n-1} + 1$. If at time t_n a cell loses the label \mathfrak{c}' due to a resampling event then $U_n = U_{n-1} - 1$. To have $U_m < N/5$ for some m satisfying $0 \leq m \leq n$ we will need $\max_{0 \leq m \leq n} |U_m - U_0| \geq N/20$. It follows from the reflection principle that there exists a constant C such that $E[\max_{0 \leq m \leq n} |U_m - U_0|] \leq C\sqrt{n}$ for all $n \geq 0$. By Markov's inequality,

$$P\left(\max_{0 \leq m \leq n} |U_m - U_0| \geq N/20\right) \leq C\sqrt{n}/N$$

for some constant C .

Let R be the number of resampling events that occur in the time interval $[T, T + \frac{1}{2}f(N)]$ that involve pairs of cells such that one is labeled \mathfrak{c}' and the other is not. By the bound on the rate at which resampling events occur of $N/4$ and Lemma 2 we have

$$P(R > k) \leq \sum_{i=k+1}^{\infty} \frac{(Nf(N))^i e^{-Nf(N)/8}}{8^i i!} \leq \frac{(Nf(N))^k}{8^k k!}.$$

Then

$$\begin{aligned} P((A'_3)^C) &\leq P(\{\max_{0 \leq m \leq R} |U_m - U_0| \geq N/20\} \cap \{R \leq N^{3/2}\}) \\ &\quad + P(\{\max_{0 \leq m \leq R} |U_m - U_0| \geq N/20\} \cap \{R > N^{3/2}\}) \\ &\leq P(\{\max_{0 \leq m \leq N^{3/2}} |U_m - U_0| \geq N/20\}) + P(R > N^{3/2}) \\ &\leq \frac{C}{N^{1/4}} + \frac{(Nf(N))^{N^{3/2}}}{8^{N^{3/2}} \lceil N^{3/2} \rceil!} \\ &\rightarrow 0 \text{ as } N \rightarrow \infty. \end{aligned}$$

Let A'_4 be the event that $\mathfrak{A}'_t = 0$ for some time $t \in [T, T + \frac{1}{2}f(N)]$. Notice that if $\mathfrak{A}'_t = 0$ then $\mathfrak{A}'_s = 0$ for $s \geq t$. Therefore, A'_4 is the event that the label \mathfrak{a}' is eliminated by time $T + \frac{1}{2}f(N)$. By the given dynamics \mathfrak{A}'_t can only increase when cells marked \mathfrak{a}' replace cells marked \mathfrak{b}' or \mathfrak{c}' via resampling events. At time t the rate at which this happens is

$$\frac{1}{2} \cdot \frac{\mathfrak{A}'_t(N - \mathfrak{A}'_t)}{N} \leq \mathfrak{A}'_t. \tag{2}$$

We define the event E as

$$E = (A'_1)^C \cap (A'_2)^C \cap A'_3 \cap \{T < \frac{1}{2}f(N)\}.$$

Selection will cause \mathfrak{A}' to decrease. On the event $(A'_2)^C$ all of the cells marked \mathfrak{c}' will have fitness at least $\frac{1}{32}W_0$ greater than any cell marked \mathfrak{a} until time $f(N)$. Thus, on the event $(A'_2)^C \cap \{T < \frac{1}{2}f(N)\}$ all of the cells marked \mathfrak{c}' will have fitness at least $\frac{1}{32}W_0$ greater than any cell marked \mathfrak{a} for all times $t \in [T, T + \frac{1}{2}f(N)]$. On the event A'_3 there are at least $N/5$ cells marked \mathfrak{c} for all times $t \in [T, T + \frac{1}{2}f(N)]$. Hence, on the event E cells marked \mathfrak{a}' will become cells marked \mathfrak{c}' by a rate of at least

$$\frac{\gamma \mathfrak{A}'_t \mathfrak{C}'_t W_0}{32N} \geq \frac{\gamma}{160} W_0 \mathfrak{A}'_t \quad (3)$$

for all times $t \in [T, T + \frac{1}{2}f(N)]$.

Let $\{U'_n\}$ be a biased random walk which goes up with probability

$$p' = \frac{160}{160 + \gamma W_0}$$

and down with probability $1 - p'$. Let N be large enough so that $p' < 1/2$.

Because the random walk is biased downward, the probability that the random walk visits a state $j < U'_0$ is 1. Once the random walk is in state j , it goes up 1 with probability p' and will eventually return to j with probability 1. The random walk will go down 1 with probability $1 - p'$ and, from basic martingale arguments, the probability that it never returns to j again is $(1 - 2p')/(1 - p')$. Therefore, once U' is in state j , the probability it never returns to state j is

$$\frac{(1 - 2p')}{1 - p'} \cdot (1 - p') = 1 - 2p'.$$

Hence the number of times U' visits a state $j < U'_0$ has the geometric distribution with mean $1/(1 - 2p')$. For more details see pages 194-196 of [3]. By equations (2) and (3) we see that on the event E , if \mathfrak{A}' changes during the time interval $[T, T + \frac{1}{2}f(N)]$ it decreases with probability higher than p' . The expected number of times that \mathfrak{A}' will visit state j is therefore less than or equal to $1/(1 - 2p')$ for any $j \in \{1, 2, \dots, N-1\}$. Also, the rate at which \mathfrak{A}'_t changes state is at least

$$\frac{\gamma}{160} W_0 \mathfrak{A}'_t$$

for all times $t \in [T, T + \frac{1}{2}f(N)]$ by equation (3). Let $\bar{A} = \{t \geq T : \mathfrak{A}'_t > 0\}$ and let λ be Lebesgue measure. Then

$$E[\lambda(\bar{A})1_E] \leq \frac{160}{(1 - 2p')\gamma W_0} \sum_{j=1}^N \frac{1}{j} \sim \frac{160 \log N}{\gamma W_0}$$

as $N \rightarrow \infty$.

Observe that

$$\begin{aligned} P(E \cap (A'_4)^C) &= P\left(E \cap \left\{\lambda(\bar{A}) \geq \frac{1}{2}f(N)\right\}\right) \\ &= P\left(\lambda(\bar{A})1_E \geq \frac{1}{2}f(N)\right) \\ &\leq \frac{2E[\lambda(\bar{A})1_E]}{f(N)} \text{ by Markov's Inequality} \\ &\rightarrow 0 \text{ as } N \rightarrow \infty. \end{aligned}$$

Therefore,

$$P(E \cap A'_4) - P\left(T < \frac{1}{2}f(N)\right) \rightarrow 0 \text{ as } N \rightarrow \infty.$$

This allows us to do the following computation:

$$\begin{aligned} 1 &= \lim_{N \rightarrow \infty} \left(P\left(T < \frac{1}{2}f(N)\right) + P\left(T \geq \frac{1}{2}f(N)\right) \right) \\ &= \lim_{N \rightarrow \infty} \left(P(E \cap A'_4) + P\left(T \geq \frac{1}{2}f(N)\right) \right) \\ &= \lim_{N \rightarrow \infty} \left(P\left((A'_1)^C \cap (A'_2)^C \cap A'_3 \cap A'_4 \cap \left\{T < \frac{1}{2}f(N)\right\}\right) + P\left(T \geq \frac{1}{2}f(N)\right) \right) \\ &\leq \lim_{N \rightarrow \infty} \left(P\left(\left\{B_T \leq \frac{1}{2}f(N)\right\} \cap \left\{T < \frac{1}{2}f(N)\right\}\right) + P\left(T \geq \frac{1}{2}f(N)\right) \right) \\ &= \lim_{N \rightarrow \infty} P\left(B_T 1_{\{T < \frac{1}{2}f(N)\}} \leq \frac{1}{2}f(N)\right). \end{aligned}$$

□

Let $B = \inf\{t : X_t^- > X_0^+ - W_0/4\}$.

Proposition 12. *Consider any sequence of initial configurations X_0 such that $W_0 \geq \log N$ for all N . As N tends to infinity,*

$$P(B > f(N)) \rightarrow 0.$$

Proof. First note that if $\mathfrak{B}_0 + \mathfrak{C}_0 \geq N/4$ then, because all of the cells labeled \mathfrak{b} or \mathfrak{c} at time 0 are also labeled \mathfrak{c}' , we have that $\mathfrak{C}'_0 \geq N/4$. The result then follows by Lemma 11 with $T = 0$. On the other hand, if $\mathfrak{B}_0 + \mathfrak{C}_0 < N/4$ then $\mathfrak{A}_0 \geq 3N/4$.

Let $T = (\inf\{t : \mathfrak{A}_t < N/4\}) \wedge (\inf\{t : \mathfrak{C}_t \geq N/4\})$. Let A_5 be the event that $\mathfrak{A}_t \geq N/4$ for all times $t \in [0, \frac{1}{2}f(N)]$. Let A_6 be the event that $\mathfrak{C}_t < N/4$ for all times $t \in [0, \frac{1}{2}f(N)]$. Define ζ to be the infimum over all times such that a cell labeled \mathfrak{b} has fitness in $(-\infty, X_0^+ - \frac{5}{32}W_0)$, a cell labeled \mathfrak{c} has fitness in $(-\infty, X_0^+ - \frac{3}{32}W_0)$, or $\mathfrak{A}_t < N/4$. Note that $A_1^C \cap A_2^C \cap A_5 \subset \{\zeta \geq \frac{1}{2}f(N)\}$.

On the event $\{\zeta \geq \frac{1}{2}f(N)\}$ the rate of increase of \mathfrak{C}_t due to selection is at least

$$\frac{\gamma \mathfrak{A}_t \mathfrak{C}_t W_0}{32N} \geq \frac{1}{128} \gamma \mathfrak{C}_t W_0 \tag{4}$$

for all $t \in [0, \frac{1}{2}f(N)]$. On the other hand, because \mathfrak{C}_t can only decrease due to resampling, \mathfrak{C}_t will decrease no faster than

$$\frac{1}{2} \cdot \frac{\mathfrak{C}_t(N - \mathfrak{C}_t)}{N} \leq \mathfrak{C}_t. \tag{5}$$

Let $\{U_n\}_{n=0}^\infty$ be a biased random walk with $U_0 = 1$ which goes up with probability

$$p = \frac{\gamma W_0}{128 + \gamma W_0}$$

and down with probability $1 - p$. Let N be large enough so that $p > 1/2$. By similar reasoning as used in the proof of Lemma 11, the number of times U_n visits a state $j \geq 1$ has the geometric

distribution with mean $1/(2p - 1)$. Also, by basic martingale arguments, the probability that U_n ever reaches state 0 is

$$\frac{1-p}{p} = \frac{128}{\gamma W_0}.$$

Note that $\mathfrak{C}_0 \geq U_0$ since the cell with the highest fitness is initially labeled \mathfrak{c} . On the event $\{\zeta \geq \frac{1}{2}f(N)\}$, we see from equations (4) and (5) that if \mathfrak{C} changes during time $[0, \frac{1}{2}f(N)]$ then it increases with a probability of at least p . Therefore, the expected number of times that \mathfrak{C} visits state j is less than or equal to $1/(2p - 1)$ and the probability the \mathfrak{C}_t reaches state 0 for some time $t \in [0, \frac{1}{2}f(N)]$ is less than $128/(\gamma W_0)$. Let A_7 be the event that \mathfrak{C}_t reaches state 0 for some time $t \in [0, \frac{1}{2}f(N)]$.

By equation (4), the rate at which \mathfrak{C} changes is at least

$$\frac{1}{128}\gamma\mathfrak{C}_t W_0$$

for all times $t \in [0, \frac{1}{2}f(N)]$ on the event $\{\zeta > \frac{1}{2}f(N)\}$. Let $\bar{C} = \{t \in [0, \frac{1}{2}f(N)] : \mathfrak{C} < \frac{1}{4}N\}$ and let λ be Lebesgue measure. Then

$$\begin{aligned} E[\lambda(\bar{C})1_{\{\zeta \geq f(N)/2\}}] &= E[\lambda(\bar{C})1_{\{\zeta \geq f(N)/2\}}1_{A_7}] + E[\lambda(\bar{C})1_{\{\zeta \geq f(N)/2\}}1_{A_7^C}] \\ &\leq \frac{1}{2}f(N)P(A_7) + \frac{128}{(2p-1)\gamma W_0} \sum_{j=1}^{\lfloor N/4 \rfloor} \frac{1}{j} \\ &\sim \frac{128 \log(N/4)}{\gamma W_0}. \end{aligned}$$

By Markov's inequality

$$\begin{aligned} P(A_1^C \cap A_2^C \cap A_5 \cap A_6) &\leq P(A_1^C \cap A_2^C \cap A_5 \cap \{\lambda(\bar{C}) \geq \frac{1}{2}f(N)\}) \\ &\leq P(\{\zeta \geq \frac{1}{2}f(N)\} \cap \{\lambda(\bar{C}) \geq \frac{1}{2}f(N)\}) \\ &= P(\lambda(\bar{C})1_{\{\zeta \geq f(N)/2\}} \geq \frac{1}{2}f(N)) \\ &\leq \frac{2E[\lambda(\bar{C})1_{\{\zeta \geq f(N)/2\}}]}{f(N)} \\ &\leq \frac{256 \log(N/4)}{\sqrt{f(N)}\gamma W_0} \text{ for } N \text{ large enough} \\ &\rightarrow 0 \text{ as } N \rightarrow \infty. \end{aligned}$$

Because $P(A_1^C \cap A_2^C) \rightarrow 1$ we have $P(A_5^C \cup A_6^C) \rightarrow 1$ as $N \rightarrow \infty$.

Note that $A_5^C \cup A_6^C \subset \{T < \frac{1}{2}f(N)\}$. Therefore, $P(T < \frac{1}{2}f(N)) \rightarrow 1$ as $N \rightarrow \infty$. Let $E_2 = (A'_1)^C \cap (A'_2)^C \cap \{T < \frac{1}{2}f(N)\}$. Then $P(E_2) \rightarrow 1$ as $N \rightarrow \infty$. To show $P(B \leq f(N)) \rightarrow 1$ we can show $P(\{B \leq f(N)\} \cap E_2) \rightarrow 1$. At time T , at least $\frac{1}{4}N$ cells will be labeled either \mathfrak{b} or \mathfrak{c} . According to the labeling all of these cells are labeled \mathfrak{c}' so that at time T we have $\mathfrak{C}_T \geq \frac{1}{4}N$. By Lemma 11 we have

$$P\left(B_T 1_{\{T < \frac{1}{2}f(N)\}} \leq \frac{1}{2}f(N)\right) \rightarrow 1 \text{ as } N \rightarrow \infty.$$

Note that

$$\left\{ B_T 1_{\{T < \frac{1}{2}f(N)\}} \leq \frac{1}{2}f(N) \right\} = \left\{ B_T \leq \frac{1}{2}f(N) \right\} \cup \left\{ T \geq \frac{1}{2}f(N) \right\}.$$

Because $E_2 \subset \{T < \frac{1}{2}f(N)\}$ we have

$$\left\{ B_T 1_{\{T < \frac{1}{2}f(N)\}} \leq \frac{1}{2}f(N) \right\} \cap E_2 = \left\{ B_T \leq \frac{1}{2}f(N) \right\} \cap E_2.$$

It then follows that

$$P \left(\left\{ B_T \leq \frac{1}{2}f(N) \right\} \cap E_2 \right) \rightarrow 1 \text{ as } N \rightarrow \infty.$$

However,

$$\left\{ B_T \leq \frac{1}{2}f(N) \right\} \cap E_2 \subset \left\{ B_T \leq \frac{1}{2}f(N) \right\} \cap \left\{ T < \frac{1}{2}f(N) \right\} \subset \{B \leq f(N)\}$$

which gives the conclusion. \square

Let $V_t^1 = \{i : X_t^i > X_0^+ + W_0/4\}$ and $V_t^2 = \{i : X_t^i < X_0^- - W_0/4\}$. Define

$$F = \inf\{t : V_t^1 \cup V_t^2 \neq \emptyset\}.$$

We now want to bound the time it takes for the width to increase.

Proposition 13. *Consider any sequence of initial configurations X_0 that depend on N such that $W_0 \geq \log N$ for all N . Then*

$$\lim_{N \rightarrow \infty} P(F > \log \log W_0) = 1.$$

Proof. By Proposition 6 with $l = W_0/4$ and $t = \log \log W_0$ we have

$$\begin{aligned} P(\inf\{s : V_s^1 \neq \emptyset\} < t) &= P\left(\sup_{0 \leq s \leq t} D_s \geq l\right) \\ &\leq \frac{2N(t\mu)^l e^{(\gamma(W_0+2l)+\mu+1)t}}{(l-1)!} \\ &\rightarrow 0 \text{ as } N \rightarrow \infty. \end{aligned}$$

By Proposition 4 with $l = W_0/4$ and $t = \log \log W_0$ we have

$$\begin{aligned} P(\inf\{s : V_s^2 \neq \emptyset\} < t) &= P\left(\sup_{0 \leq s \leq t} (X_0^- - X_s^-) \geq l\right) \\ &\leq \frac{N(t\mu)^l e^t}{l!} \\ &\rightarrow 0 \text{ as } N \rightarrow \infty. \end{aligned}$$

\square

Recall that $Y_i = \sup_{s_i \leq s \leq t_{i+1}} D_s - D_{s_i}$ and that $\{\mathcal{F}_t\}_{t \geq 0}$ is the natural filtration associated with X . Note that if $W_0 < 2 \log N$ then for all $n \geq 1$ the width satisfies $W_{s_n} = \lceil 2 \log N \rceil$.

Proposition 14. Let $X_0^i = 0$ for $1 \leq i \leq N$. For N large enough we have $E[Y_i | \mathcal{F}_{s_i}] \leq 5 \log N$ for all $i \geq 1$.

Proof. We consider a sequence of initial configurations X_0 depending on N such that for all N we have $W_0 = \lceil 2 \log N \rceil$. Because $W_0 \geq 2 \log N$ we have $s_1 = 0$ and $Y_1 = \sup_{0 \leq s \leq t_2} D_s - D_0$. We will show that for N large enough, $E[Y_1] < 5 \log N$. The result then follows because X is a strong Markov process.

We make the following definitions:

$$\begin{aligned} V_t^1(s) &= \{i : X_t^i > X_s^+ + W_s/4\} \text{ for } s \geq 0 \\ V_t^2(s) &= \{i : X_t^i < X_s^- - W_s/4\} \text{ for } s \geq 0 \\ F_0 &= B_0 = r_0 = 0 \\ F_n &= \inf_{t \geq r_{n-1}} V_t^1(r_{n-1}) \cup V_t^2(r_{n-1}) \neq \emptyset \text{ for } n \geq 1 \\ B_n &= \inf_{t \geq r_{n-1}} X_t^- > X_{r_{n-1}}^+ - W_{r_{n-1}}/4 \text{ for } n \geq 1 \\ r_n &= F_n \wedge B_n \text{ for } n \geq 1 \\ n_* &= \inf\{n \geq 1 : W_{r_n} < \log N\}. \end{aligned}$$

Note that r_1 is the first time that the event $F \cup B$ occurs and that, conceptually, r_n acts like the first time that $F \cup B$ occurs when the process is started at time r_{n-1} for $n \geq 2$. The random variables F_n and B_n play the roles of the events F and B when the processes are started at time r_{n-1} .

On the event $n-1 < n_*$, by Proposition 12 and the strong Markov property of X we have $P(B_n \leq r_{n-1} + f(N) | \mathcal{F}_{r_{n-1}}) \rightarrow 1$ uniformly on a set of probability 1 as $N \rightarrow \infty$. Likewise, on the event $n-1 < n_*$, by Proposition 13 and the strong Markov property we also have $P(F_n > r_{n-1} + f(N) | \mathcal{F}_{r_{n-1}}) \rightarrow 1$ uniformly on a set of probability 1 as $N \rightarrow \infty$. Therefore, on the event $n-1 < n_*$ we have $P(B_n < F_n | \mathcal{F}_{r_{n-1}}) \rightarrow 1$ uniformly on a set of probability 1.

Because the bounds in Propositions 12 and 13 do not depend on n we can choose a sequence p_N such that $p_N \rightarrow 1$ as $N \rightarrow \infty$ and almost surely

$$p_N 1_{\{n-1 < n_*\}} \leq P(B_n < F_n | \mathcal{F}_{r_{n-1}}) 1_{\{n-1 < n_*\}}$$

for all $n \geq 0$. Let $\{S_n\}_{n=0}^\infty$ be a random walk starting at 1 which goes down 1 with probability p_N and up 1 with probability $1 - p_N$ until it reaches 0. Once S reaches 0 it is fixed. For $n < n_*$ we couple S with X as follows:

- Each step of the process S corresponds to a time r_n .
- On the event $\{F_n < B_n\}$ we have $S_n - S_{n-1} = 1$.
- On the event $\{B_n \leq F_n\}$ we have $S_n - S_{n-1} = -1$ with probability $p_N / P(B_n \leq F_n)$ and we have $S_n - S_{n-1} = -1$ with probability $1 - p_N / P(B_n \leq F_n)$.

We will show that this coupling is well-defined and gives the necessary bound. Initially, $S_0 = 1$ and $2^{S_0-1} W_0 = W_0$. On the event that $B_n \leq F_n$, we have

$$W_{r_n} < \frac{1}{2} W_{r_{n-1}} \text{ and } \sup_{r_{n-1} \leq t \leq r_n} D_t - D_{r_{n-1}} \leq \frac{1}{4} W_{r_{n-1}}.$$

On the event that $F_n < B_n$, we have

$$W_{r_n} < 2W_{r_{n-1}} \text{ and } \sup_{r_{n-1} \leq t \leq r_n} D_t - D_{r_{n-1}} \leq \frac{1}{4}W_{r_{n-1}} + 1.$$

Therefore, if $2^{S_{n-1}-1}W_0 \geq W_{r_{n-1}}$ then $2^{S_n-1}W_0 \geq W_{r_n}$ by the coupling. It follows that $2^{S_n-1}W_0 \geq \sup_{r_{n-1} \leq t \leq r_n} D_t - D_{r_{n-1}}$ as well. By induction, $2^{S_n-1}W_0 \geq W_{r_n}$ for all $n < n_* \wedge \inf\{m : S_m = 0\}$. If $n = \inf\{m : S_m = 0\}$ then $W_{r_n} \leq \log N$. Therefore $n_* \leq \inf\{m : S_m = 0\}$ and the induction holds for all $n < n_*$.

We define a function d on $(\{0\} \cup \mathbb{N})^\infty$ such that if $x = (x_0, x_1, \dots)$ then

$$d(x) = \sum_{i=0}^{\infty} 1_{\{x_i > 0\}} 2^{x_i-1} W_0.$$

Consider $S = (S_0, S_1, \dots)$ as a random element in $(\{0\} \cup \mathbb{N})^\infty$. Then

$$d((S_0, S_1, \dots, S_n, 0, 0, \dots)) \geq \sum_{i=1}^n \left(\sup_{r_{i-1} \leq t \leq r_i} D_t - D_{r_{i-1}} \right) \geq \sup_{0 \leq t \leq r_n} D_t$$

for all n such that $n-1 < n_*$. By definition, n_* is the first n such that $W_{r_n} < \log N$. Hence $d(S) \geq Y_1$.

For any $n \geq 0$ we have

$$P(S_{2n+1} = 0) = \binom{2n+1}{n} (1-p_N)^n p_N^{n+1} \leq 4^n (1-p_N)^n p_N^{n+1}.$$

If $S_{2n+1} = 0$ then

$$d(S) \leq \left(2 + 2 \sum_{i=1}^n 2^{i-1} \right) W_0 = 2^{n+1} W_0$$

which is obtained by taking n steps up followed by $n+1$ steps down.

Therefore,

$$E[Y_1] \leq E[d(S)] \leq \sum_{n=0}^{\infty} [4(1-p_N)]^n p_N^{n+1} 2^{n+1} W_0 = \frac{2p_N W_0}{1 - 8(1-p_N)p_N} \sim 4 \log N$$

because $W_0 = \lceil 2 \log N \rceil$ and $p_N \rightarrow 1$ as $N \rightarrow \infty$. This shows that for N large enough we have $E[Y_1] < 5 \log N$, which gives the conclusion. \square

5 Bounding the frequency of getting large widths

For this section let $l = \lfloor \log N \rfloor$ and $t = \frac{1}{4\gamma} \log \log N$. We define

$$K_1 = \frac{2N(t\mu)^l e^{(\gamma(W_0+2l)+\mu+1)t}}{(l-1)!} \text{ and}$$

$$K_2 = \frac{N(t\mu)^l e^t}{l!}.$$

Lemma 15. Consider any sequence of initial configurations X_0 that depend on N such that $W_0 \leq \log N$. Then

$$P \left(\sup_{0 \leq s \leq t} W_s \leq 2 \log N \right) \geq 1 - K_1 - K_2.$$

Proof. By Proposition 6 we have

$$P \left(\sup_{0 \leq s \leq t} D_s \geq l \right) \leq K_1.$$

By Corollary 4 we have

$$P \left(\sup_{0 \leq s \leq t} (X_0^- - X_s^-) \geq l \right) \leq K_2.$$

If $\sup_{0 \leq s \leq t} D_s \leq \log N$ and $\sup_{0 \leq s \leq t} X_0^- - X_s^- \leq \log N$ then $\sup_{0 \leq s \leq t} W_s \leq 2 \log N$. This gives the result. \square

Proposition 16. Let $W_0 = 0$. For N large enough,

$$\sup_{s \in [1, \infty)} \frac{1}{s} E[N_s] \leq \frac{1}{t}$$

Proof. Notice that

$$\{N_s \geq i\} = \{s_i \leq s\} \subset \left\{ \sum_{j=1}^i (s_j - t_j) \leq s \right\}.$$

Therefore,

$$P(N_s \geq i) \leq P \left(\sum_{j=1}^i (s_j - t_j) \leq s \right).$$

Applying Lemma 15 and the strong Markov property of X we have

$$P(s_j - t_j \geq t | \mathcal{F}_{t_j}) \geq 1 - K_1 - K_2$$

for all j . Taking expectations of both sides yields

$$\begin{aligned} 1 - K_1 - K_2 &\leq E \left[\inf_j P(s_j - t_j \geq t | \mathcal{F}_{t_j}) \right] \\ &\leq \inf_j E \left[P(s_j - t_j \geq t | \mathcal{F}_{t_j}) \right] \\ &= \inf_j P(s_j - t_j \geq t). \end{aligned}$$

Let $p_N = 1 - K_1 - K_2$. Note that $p_N \rightarrow 1$ as $N \rightarrow \infty$. Define an i.i.d. sequence $\{V_i\}_{i=1}^\infty$ of random variables with distribution $P(V_i = 0) = 1 - p_N$ and $P(V_i = t) = p_N$. Then

$$P \left(\sum_{j=1}^i (s_j - t_j) \leq s \right) \leq P \left(\sum_{j=1}^i V_i \leq s \right).$$

This will allow us to define a new process N'_s such that $N'_s = i$ if

$$\sum_{j=1}^i V_j \leq s < \sum_{j=1}^{i+1} V_j.$$

Note that $P(N'_s = 0) = p_N$ for $s \in [0, t)$ and that $P(N'_s \geq k) \geq P(N_s \geq k)$ for all k . Therefore, it is enough to bound $E[N'_s]/s$.

Let $V_0 = 0$. Jumps of the process N'_s only occur at points kt where k is a positive integer. On the time interval $[1, t)$ the process N'_s is constant and has value $\max\{i \geq 0 : V_i = 0\}$. Therefore N'_s has the shifted geometric distribution for $s \in [0, t)$ with mean $(1 - p_N)/p_N$. We can now make use of the fact that N'_s is a Markov process. If we consider values at kt for $k \geq 0$ we have for $s \in [(k-1)t, kt)$ that $E[N'_s] = k(1 - p_N)/p_N$. For $k \geq 2$ we then have

$$\frac{1}{s} E[N'_s] = \frac{k(1 - p_N)}{sp_N} \leq \frac{k(1 - p_N)}{(k-1)p_N t}.$$

This gives us

$$\frac{t}{s} E[N'_s] \leq \frac{k(1 - p_N)}{(k-1)p_N} \rightarrow 0 \text{ as } N \rightarrow \infty.$$

On the time interval $[1, t)$ we have

$$\frac{1}{s} E[N'_s] \leq \frac{(1 - p_N)}{p_N} \rightarrow 0 \text{ as } N \rightarrow \infty.$$

□

6 Proof of Theorem 1

Proof of Theorem 1. Fix $t > 1$ and $\epsilon > 0$. There exists N_0 which does not depend on t such that for any $N \geq N_0$ we have

$$\begin{aligned} E\left[\frac{D_t}{t}\right] &\leq E\left[\frac{D'_t + \sum_{i=1}^{N_t} Y_i}{t}\right] \text{ by Proposition 9} \\ &= E\left[\frac{D'_t}{t}\right] + E\left[\frac{\sum_{i=1}^{N_t} Y_i}{t}\right] \\ &\leq (1 + 2\gamma + \epsilon)\bar{k} + \frac{1}{t} E\left[\sum_{i=1}^{N_t} Y_i\right] \text{ by Proposition 8} \\ &\leq (1 + 2\gamma + \epsilon)\bar{k} + \frac{1}{t} E\left[\sum_{i=1}^{N_t} Y_i\right] \\ &= (1 + 2\gamma + \epsilon)\bar{k} + \frac{1}{t} \sum_{i=1}^{\infty} E[Y_i 1_{\{N_t \geq i\}}] \\ &= (1 + 2\gamma + \epsilon)\bar{k} + \frac{1}{t} \sum_{i=1}^{\infty} E[E[Y_i 1_{\{N_t \geq i\}} | \mathcal{F}_{s_i}]] \end{aligned}$$

$$\begin{aligned}
&= (1 + 2\gamma + \epsilon)\bar{k} + \frac{1}{t} \sum_{i=1}^{\infty} E[1_{\{N_t \geq i\}} E[Y_i | \mathcal{F}_{s_i}]] \\
&\leq (1 + 2\gamma + \epsilon)\bar{k} + \frac{5 \log N}{t} \sum_{i=1}^{\infty} E[1_{\{N_t \geq i\}}] \text{ by Proposition 14} \\
&= (1 + 2\gamma + \epsilon)\bar{k} + \frac{5 \log N}{t} E[N_t].
\end{aligned}$$

For N large enough, $\bar{k} \leq 2 \log N / \log \log N$. Therefore, for N large enough,

$$\begin{aligned}
(1 + 2\gamma + \epsilon)\bar{k} + \frac{5 \log N}{t} E[N_t] &\leq 2(1 + 2\gamma + \epsilon) \frac{\log N}{\log \log N} + \frac{5 \log N}{t} E[N_t] \\
&\leq 2(1 + 2\gamma + \epsilon) \frac{\log N}{\log \log N} + \frac{20\gamma \log N}{\log \log N} \text{ by Proposition 16} \\
&= 2(1 + 12\gamma + \epsilon) \frac{\log N}{\log \log N}.
\end{aligned}$$

□

7 Acknowledgements

I would like to thank Jason Schweinsberg for suggesting the problem, patiently helping me work through various parts of the proof and for helping me revise the first drafts of the paper.

References

- [1] E. Brunet, I. Rouzine, C. Wilke (2008). The stochastic edge in adaptive evolution. *Genetics* **179**, 603-620.
- [2] M. Desai, D.S. Fisher, (2007). Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics* **176**, 1759-1798.
- [3] R. Durrett. Probability Models for DNA Sequence Evolution, 2nd Ed. Springer, 2008.
- [4] R. Durrett and J. Mayberry (2011). Travelling waves of selective sweeps. *Ann. Appl. Probab.* **21**, 699-744.
- [5] A. Etheridge and F. Yu (2011). Girsanov Transformation and the Rate of Adaptation. Preprint.
- [6] J. Haigh (1978). The accumulation of deleterious genes in a population. *Theor. Popul. Biol.* **14**, 251-267.
- [7] W.G. Hill and A. Robertson (1966). The effect of linkage on limits to artificial selection. *Genetics Research* **8**, 269-294.
- [8] S.C. Park, D. Simon, J. Krug (2010). The speed of evolution in large asexual populations. *J. Stat. Phys.* **138**, 381-410.

- [9] I. Rouzine, E. Brunet, C. Wilke (2007). The traveling-wave approach to asexual evolution: Muller's ratchet and speed of adaptation. *Theor. Popul. Biol.* **73**, 24-46.
- [10] F. Yu, A. Etheridge (2008), Rate of Adaptation of Large Populations in *Evolutionary Biology from Concept to Application*, Springer, 3-27.
- [11] F. Yu, A. Etheridge, C. Cuthbertson (2009). Asymptotic Behaviour of the Rate of Adaptation. *Ann. Appl. Probab.* **20**, 978-1004.